Back Propagation

By Suneesh Jacob

November 2024

Problem

Suppose we have an artificial neural network with the weights and bias values as shown in the figure. For the specified input values and the expected output values shown in the figure, compute the updated weights for gradient descent method using back propagation algorithm with a learning rate of $\alpha = 0.6$.

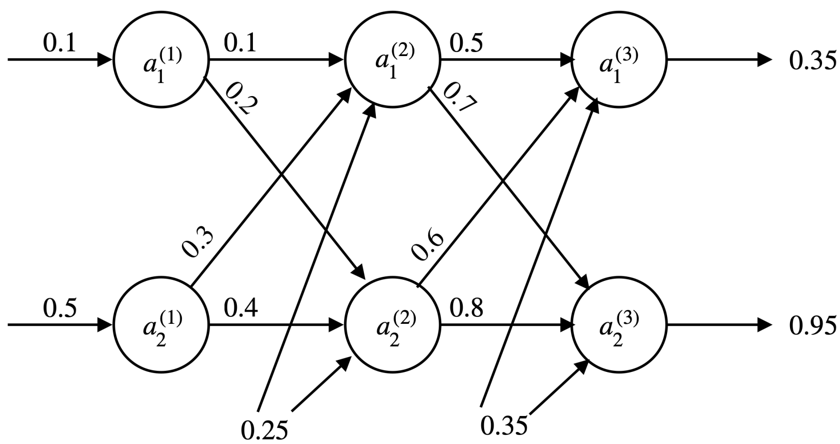

Figure 1: Artificial Neural Network

Solution

From the figure, the given data can be gathered as

$$a^{(1)} = x = \begin{bmatrix} 0.1 \\ 0.5 \end{bmatrix}$$

$$w^{(1)} = \begin{bmatrix} 0.1 & 0.3 \\ 0.2 & 0.4 \end{bmatrix}$$

$$b^{(1)} = \begin{bmatrix} 0.25 \\ 0.25 \end{bmatrix}$$

$$w^{(2)} = \begin{bmatrix} 0.5 & 0.6 \\ 0.7 & 0.8 \end{bmatrix}$$

$$b^{(2)} = \begin{bmatrix} 0.35 \\ 0.35 \end{bmatrix}$$

$$y = \begin{bmatrix} 0.05 \\ 0.95 \end{bmatrix}$$

Forward pass

For forward pass, we find $z^{(2)}$, $a^{(2)}$, $z^{(3)}$ and $a^{(3)}$ in the same sequence.

$$z^{(2)} = w^{(1)}a^{(1)} + b^{(1)} = \begin{bmatrix} 0.41 \\ 0.47 \end{bmatrix}$$

$$a^{(2)} = \sigma\left(z^{(2)}\right) = \begin{bmatrix} 0.60 \\ 0.62 \end{bmatrix}$$

$$z^{(3)} = w^{(2)}a^{(2)} + b^{(2)} = \begin{bmatrix} 1.02 \\ 1.26 \end{bmatrix}$$

$$a^{(3)} = \sigma\left(z^{(3)}\right) = \begin{bmatrix} 0.73 \\ 0.78 \end{bmatrix}$$

Back propagation

Some preliminary calculations for back propagation

In order to compute back propagation for this problem, we need to find $\sigma'\left(z^{(3)}\right)$, $\sigma'\left(z^{(2)}\right)$ and $\delta^{(3)}$.

Firstly, we know that $\sigma'(x) = \sigma(x)(1 - \sigma(x))$. By using this, $\sigma'\left(z^{(3)}\right)$ and $\sigma'\left(z^{(2)}\right)$ can be found as shown below.

$$\sigma'\left(z^{(3)}\right) = \sigma'\left(\begin{bmatrix} 1.02 \\ 1.26 \end{bmatrix}\right) = \begin{bmatrix} \sigma(1.02)(1 - \sigma(1.02)) \\ \sigma(1.26)(1 - \sigma(1.26)) \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.17 \end{bmatrix}$$

$$\sigma'\left(z^{(2)}\right) = \sigma'\left(\begin{bmatrix} 0.41 \\ 0.47 \end{bmatrix}\right) = \begin{bmatrix} \sigma(0.41)(1 - \sigma(0.41)) \\ \sigma(0.47)(1 - \sigma(0.47)) \end{bmatrix} = \begin{bmatrix} 0.24 \\ 0.24 \end{bmatrix}$$

Now, we have the loss function to be

$$L = \frac{1}{2}\left(\frac{\left(a_1^{(3)} - y_1\right)^2 + \left(a_2^{(3)} - y_2\right)^2}{2}\right)$$

$$\delta^{(3)} = \frac{\partial L}{\partial a^{(3)}} = \frac{1}{2}\begin{bmatrix} a_1^{(3)} - y_1 \\ a_2^{(3)} - y_2 \end{bmatrix} = \begin{bmatrix} 0.68 \\ -0.17 \end{bmatrix}$$

Back propagation

For backward pass, we find $\frac{\partial L}{\partial b^{(2)}}$, $\frac{\partial L}{\partial w^{(2)}}$, $\delta^{(2)}$, $\frac{\partial L}{\partial b^{(1)}}$ and $\frac{\partial L}{\partial w^{(1)}}$ in the same sequence.

$$\frac{\partial L}{\partial b^{(2)}} = \delta^{(3)} \odot \sigma'\left(z^{(3)}\right) = \begin{bmatrix} 0.68 \\ -0.17 \end{bmatrix} \odot \begin{bmatrix} 0.2 \\ 0.17 \end{bmatrix} = \begin{bmatrix} 0.68 \times 0.2 \\ -0.17 \times 0.17 \end{bmatrix} = \begin{bmatrix} 0.13 \\ -0.03 \end{bmatrix}$$

$$\frac{\partial L}{\partial w^{(2)}} = \frac{\partial L}{\partial b^{(2)}}\left(a^{(2)}\right)^T = \begin{bmatrix} 0.13 \\ -0.03 \end{bmatrix}\begin{bmatrix} 0.60 & 0.62 \end{bmatrix} = \begin{bmatrix} 0.08 & 0.08 \\ -0.02 & -0.02 \end{bmatrix}$$

$$\delta^{(2)} = \left(w^{(2)}\right)^T \frac{\partial L}{\partial b^{(2)}} = \begin{bmatrix} 0.5 & 0.7 \\ 0.6 & 0.8 \end{bmatrix}\begin{bmatrix} 0.13 \\ -0.03 \end{bmatrix} = \begin{bmatrix} 0.05 \\ 0.06 \end{bmatrix}$$

$$\frac{\partial L}{\partial b^{(1)}} = \delta^{(2)} \odot \sigma'\left(z^{(2)}\right) = \begin{bmatrix} 0.05 \\ 0.06 \end{bmatrix} \odot \begin{bmatrix} 0.24 \\ 0.24 \end{bmatrix} = \begin{bmatrix} 0.05 \times 0.24 \\ 0.06 \times 0.24 \end{bmatrix} = \begin{bmatrix} 0.01 \\ 0.01 \end{bmatrix}$$

$$\frac{\partial L}{\partial w^{(1)}} = \frac{\partial L}{\partial b^{(1)}}\left(a^{(1)}\right)^T = \begin{bmatrix} 0.01 \\ 0.01 \end{bmatrix}\begin{bmatrix} 0.60 & 0.62 \end{bmatrix} = \begin{bmatrix} 0.08 & 0.08 \\ 0.1 & 0.5 \end{bmatrix}$$

Updating the weights

For $\alpha = 0.6$, the new weights and biases would be

$$w_{\text{new}}^{(1)} = w^{(1)} - \alpha\frac{\partial L}{\partial w^{(1)}} = \begin{bmatrix} 0.1 & 0.3 \\ 0.2 & 0.4 \end{bmatrix} - 0.6\begin{bmatrix} 0.08 & 0.08 \\ 0.1 & 0.5 \end{bmatrix} = \begin{bmatrix} 0.1 & 0.3 \\ 0.2 & 0.4 \end{bmatrix}$$

$$b_{\text{new}}^{(1)} = b^{(1)} - \alpha\frac{\partial L}{\partial b^{(1)}} = \begin{bmatrix} 0.24 \\ 0.24 \end{bmatrix}$$

$$w_{\text{new}}^{(2)} = w^{(2)} - \alpha\frac{\partial L}{\partial w^{(2)}} = \begin{bmatrix} 0.45 & 0.55 \\ 0.71 & 0.81 \end{bmatrix}$$

$$b_{\text{new}}^{(2)} = b^{(2)} - \alpha\frac{\partial L}{\partial b^{(2)}} = \begin{bmatrix} 0.27 \\ 0.37 \end{bmatrix}$$

Note 1: Here, the loss function is considered to be $L = \frac{1}{2}\left(\frac{\left(a_1^{(3)} - y_1\right)^2 + \left(a_2^{(3)} - y_2\right)^2}{2}\right)$, but in some books it may be considered to be just $L = \frac{1}{2}\left(\left(a_1^{(3)} - y_1\right)^2 + \left(a_2^{(3)} - y_2\right)^2\right)$, in which case the updated weights may differ at each iteration, however in both cases the same weights are expected at the end of the gradient descent algorithm.
Note 2: The values shown in this document are rounded up to 2 decimal points.